Homework 6 – STAT 521

Due on Nov. 25, 2015

1. Consider the linear model

$$Y = X\theta + \varepsilon, E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 \Sigma,$$

where X is of order $n \times m$, and Σ is a symmetric $n \times n$ matrix with $rank(\Sigma) = k < n$ (not positive definite).

(1). Show that there are $(n-k)$ linearly independent vectors $p_1, ..., p_{n-k} \in R^n$ such that $p_i' Y = p_i' X\theta$ with probability 1 for any $i = 1, ..., n-k$.

(2). Show that there exists a $n \times k$ matrix A of rank k, such that $Z = A'Y$ follows linear model

$$Z = A'X\theta + \varepsilon^*, E(\varepsilon^*) = 0, Var(\varepsilon^*) = \sigma^2 I_k.$$

2. (**Auditor Training Example**) An accounting firm tested three training methods in statistical sampling: (1). study at home with training materials; (2). training at local office by local staff; (3). training in headquarter office at Chicago by national staff. The auditors were grouped into 4 blocks according to time elapsed since college graduation, and auditors in each block were assigned at random to the three training methods.A proficiency measure based on the final analysis was obtained:

Training Methods	Block (j)			
(i)	1	2	3	4
1	73	76	75	76
2	81	78	76	71
3	92	89	87	88

(1). Use the two-way ANOVA model (randomized block model):

$$Y_{ij} = \tau_i + \beta_j + \varepsilon_{ij}, i = 1, ..., 3; j = 1, ..., 4$$

define $\tau = (\tau_1, \tau_2, \tau_3), \beta = (\beta_1, \beta_2, \beta_3, \beta_4)$. Specify matrix C and Q in the reduced normal equation $C\tau = Q$.

(2). Find the BLUEs for treatment (training) difference $\tau_i - \tau_{i'}$, for any $i < i'$.

(3). Testing hypothesis that treatment have equal effects, $H_0 : \tau_{1,} = \tau_2 = \tau_3$.

3. (One-way ANOVA) Suppose the linear regression model is given by

$$Y_{ij} = \mu_i + \varepsilon_{ij}, i = 1, ...k; j = 1, ..., n$$

where $\varepsilon_{ij}^{\prime}s$ are i.i.d errors following $N\left(0, \sigma^2\right)$. In the matrix form, response vector is expressed as

$$Y = \left(Y_{11}, ...Y_{1n}; ...; Y_{k1}, ...Y_{kn}\right)'$$

and parameter vector is $\mu = \left(\mu_1, ..., \mu_k\right)'$.

(1). Find the design matrix in the model $Y = X\ \mu + \varepsilon$ and determine its rank.

(2). Based on $R_0^2 = \left(Y - X\hat{\mu}\right)'\left(Y - X\hat{\mu}\right)$, where $\hat{\mu}$ is the least square estimator, find an unbaised estimator for σ^2.

(3). Consider null hypothesis $H_0 : \mu_1 = ... = \mu_k$. Under $\mathrm{H}_{0,}$ rewrite the design matrix X as X_0, determine

$$R_1^2 = \min_{\mu \in R}\left(Y - X_0\mu\right)'\left(Y - X_0\mu\right).$$

(4). Show that

$$R_1^2 - R_0^2 = n\sum_{i=1}^{k}\left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2,$$

where $\bar{Y}_{i\cdot} = \frac{1}{n}\sum\limits_{i=1}^{k} Y_{ij}, \bar{Y}_{\cdot\cdot} = \frac{1}{nk}\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n} Y_{ij}$.

(5). Rewrite the null hypothesis in (3) as $H_0 : H\mu = \mathbf{0}$, where matrix H is a $(k-1) \times k$ matrix and $M(H) \subset M(X')$. Find the matrix A such that $Var\left(H\hat{\mu}\right) = \sigma^2 A$, with the estimator $\hat{\mu}$ in (2). Show that

$$\left(H\hat{\mu}\right)' A^{-1}\left(H\hat{\mu}\right) = n\sum_{i=1}^{k}\left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2.$$